Integrating Multiple Models for Enhanced Electricity Demand Forecasting Considering Multifactor Influences

Sangbing Tsai[1*], Jaheer Mukthar.KP [2]

[1]International Engineering and Technology Institute, Hong Kong; klj0418@gmail.com

[2]Kristu Jayanti College Autonomous Bengaluru, India; jaheermukthar@gmail.com

*Corresponding Author: klj0418@gmail.com

DOI: https://doi.org/10.30209/IJMO.202503.006

Submitted: Feb. 16, 2025 Accepted: Apr. 08, 2025

ABSTRACT

Electricity demand forecasting plays a crucial role in the planning and operation of power systems. This study aims to enhance the accuracy and generalization capability of electricity demand forecasting to improve the efficiency of electricity production, distribution, and utilization. In this research, a high-performance electricity demand forecasting model is constructed by introducing innovative designs, including ConvTrans models, attention mechanism, and the word2vec module. Through comparison experiments on multiple public datasets, our model, when compared to six baseline models, demonstrates significant improvements in performance metrics such as MAE, R^2, Pearson Correlation Coefficient, and F1-score. These results validate the superior performance of the proposed model in electricity demand forecasting tasks and affirm its high robustness and generalization capabilities.

Keywords: Transformer, Attention mechanism, Electricity demand forecasting, Power systems

1 Introduction

Electricity demand forecasting [1] plays a crucial role in the planning and operation of power systems, with its accuracy directly impacting the effective utilization of electrical energy and the stable operation of the system [2, 3]. Firstly, accurate electricity demand forecasting optimizes the scheduling and operation of power systems, ensuring an adequate power supply to meet the continuously growing demand. This not only contributes to enhancing the reliability of power systems but also reduces excessive reliance on backup power generation units, lowering operational costs, and improving overall efficiency. Secondly, precise electricity demand forecasting is essential for the integration and management of renewable energy sources such as wind and solar power [4]. Renewable energy sources exhibit characteristics of instability and intermittency, with their generation influenced by various factors, including weather conditions. By integrating weather data, quarters, holidays, and other influencing factors, improving the accuracy of electricity demand forecasting enables better adaptation to the volatility of renewable energy sources [5]. Additionally, through accurately predicting electricity demand, it becomes possible to better coordinate the combination of traditional and clean energy sources, minimizing reliance on high-carbon energy sources [6]. This contributes to achieving carbon neutrality goals for power systems [7], mitigating the impacts of climate change, and advancing the development of green and sustainable energy [8].

The application of deep learning techniques in electricity demand forecasting [9] significantly enhances the performance and accuracy of predictive models, providing more reliable decision support for power system planning and operation. Deep learning , with its powerful feature learning and pattern recognition capabilities [10], demonstrates immense potential in the field of electricity demand forecasting. Firstly, deep learning models can automatically extract complex non-linear features from extensive time-series electricity data [11], capturing latent factors influencing electricity demand. This end-to-end learning approach helps circumvent the laborious process of manual feature extraction, allowing models to better adapt to the dynamic changes in power systems. Secondly, the widespread use of deep learning techniques such as recurrent neural networks (RNNs) [12] and long short-term memory networks (LSTMs) [13] enables models to capture long-term dependencies within time series data. This is crucial for data with temporal characteristics, effectively addressing the challenges of capturing long-term influences that traditional methods struggle with. Additionally, deep learning structures such as convolutional neural networks (CNNs) [14] exhibit significant advantages in handling spatial information related to electricity demand, such as the impact of geographical locations. The introduction of deep learning enables models to consider multiple dimensions of influencing factors comprehensively, improving the global and holistic nature of predictions. Deep learning models also demonstrate superior performance in handling multi-source heterogeneous data [15], integrating information from various dimensions such as weather data, socio-economic indicators, and seasonal factors to capture the complex relationships influencing electricity demand more accurately. Moreover, attention mechanisms [16] within deep learning technologies allow models to focus more on crucial time points or specific factors contributing significantly to electricity demand, enhancing the adaptability of the model across different time scales. The widespread application of deep learning technologies in electricity demand forecasting also holds promise for achieving intelligence and adaptability within power systems. Through end-to-end learning on large-scale, high-dimensional data, deep learning models continually optimize their structures, gradually adapting to new patterns and changes in power system operations. This provides a more flexible and advanced tool for future planning and management of power systems. In summary, the application of deep learning techniques in electricity demand forecasting brings significant significance in improving prediction accuracy, adapting to the complex and dynamic environment of power systems, and comprehensively considering multi-source information. Its introduction not only propels advancements in research and practices in electricity demand forecasting but also lays a solid foundation for constructing more intelligent and efficient power systems. Five deep learning models commonly used in the field of electricity demand forecasting include:

1. AutoRegressive Integrated Moving Average (ARIMA) [17]: ARIMA utilizes autoregressive and moving average components in time series analysis to forecast future electricity demand based on historical data [18]. Its advantage lies in its suitability for short-term predictions, excellent performance in capturing seasonality and periodicity, and its simple and interpretable model structure.

2. Support Vector Machine Regression (SVR) [19]: SVR employs a kernel function for non-linear mapping, seeking an optimal hyperplane for non-linear regression, making it suitable for capturing complex non-linear relationships in electricity demand. Its strengths include robustness against outliers, strong generalization performance, and applicability to high-dimensional data.

3. Random Forest Regression [20]: Random Forest combines multiple decision trees to make predictions, utilizing voting to decide the result. It is well-suited for capturing various factors influencing electricity demand. Its advantages include high accuracy and robustness to noise.

4. Recurrent Neural Network (RNN) [21]: RNN captures long-term dependencies in time series by leveraging a recurrent structure, making it suitable for modeling dynamic changes in electricity demand. Its advantages lie in its sensitivity to temporal patterns and its ability to capture complex historical patterns.

5. Convolutional Neural Network (CNN) [22]: CNN processes spatial correlations through convolutional operations, particularly useful for considering the impact of factors like geographical location on electricity demand. Its advantages include expertise in handling spatially rich data, making it suitable for multidimensional information processing.

This study collected data encompassing various influencing factors, including historical demand data, weather information, quarters, holidays, and dates. These factors were integrated into a single dataset, and the temporal alignment of information was performed. Subsequently, for each influencing factor, Word2Vec models [23] were employed to embed weather, quarters, holidays, and months separately, learning distributed representations for each factor. The historical load data was treated as a time series, employing a sliding window method [24] to generate Word Embeddings for the temporal data. In the subsequent Encoder layers, the encoder structure of the Convolution Transformer model (ConvTrans) [25, 26] was utilized to process all embedded sequences. Multiple layers of ConvTrans model were then employed to handle information at different abstraction levels. Simultaneously, an Attention mechanism was incorporated, allowing the model to attend to complex relationships between various influencing factors, facilitating a more comprehensive capture of their impact on the changes in electricity demand.

The three main contributions of this study are as follows:

1. Integration and Alignment of Multifactor Data: This study integrated various key factors, including historical load, weather, quarters, holidays, and months, into a comprehensive dataset. Through temporal alignment of this data, consistency and comparability within the model were ensured, enhancing predictive accuracy.

2. Word2Vec-Based Multifactor Embedding: The study employed Word2Vec models to embed each influencing factor, learning distributed representations for each. This approach enables the model to better comprehend and utilize the semantic information of each factor, thereby improving its ability to model complex relationships.

3. Multilayer ConvTrans Structure and Attention Mechanism: The study introduced an encoder structure to process embedded sequences through multiple layers of ConvTrans model, addressing information at various abstraction levels. The incorporation of an Attention mechanism allows the model to comprehensively capture intricate relationships among different factors, enhancing its nuanced understanding and predictive capabilities regarding changes in electricity demand.

The structure of this paper is outlined as follows. Firstly, Section 1, the introduction, provides the background and motivation for the research. Section 2 conducts an extensive literature review, presenting existing knowledge related to the issue of electricity demand forecasting. Section 3, the methodology, elaborates on the data collection process, the design principles of predictive models, including the innovative integrated Word2Vec, attention mechanism and ConvTrans layers. Section 4, the results, presents the outcomes of simulation experiments using some public datasets and evaluates the performance of the models, with an accompanying analysis of the experimental results. Section 5, the discussion, summarizes the research, encapsulates the key findings again, underscores the contributions of this study, and suggests future research directions based on identified gaps in the study.

2 Related works

2.1 Word2vec Embedding Method

The Word2Vec embedding technique is a neural network-based word embedding method designed to map words to vectors in a high-dimensional space, where semantically similar words are represented by vectors that are closer in space. The core idea behind this technology is to learn the semantic representation of words by leveraging contextual information, ensuring that words with similar meanings have analogous representations in the vector space. One application of Word2Vec in the field of power systems involves embedding various influencing factors, such as weather, seasons, etc. By representing these factors as Word2Vec vectors, the model gains a better understanding of the semantic relationships between them, thereby more accurately capturing their impact on electricity demand. Another application lies in the embedding of historical load data. By embedding historical load data into the Word2Vec vector space, the model acquires a semantic representation of load data, enhancing its understanding of load trends and periodic variations. This forms a crucial information foundation for constructing more precise electricity demand forecasting models. Additionally, Word2Vec technology can be employed for the integration of multimodal information. Given the diverse and multi-sourced nature of information in power systems, including textual descriptions and numerical data, Word2Vec facilitates the unification of these different data types into a single vector space. This enables the model to have a more comprehensive and consistent information perspective, aiding in better understanding and utilization of various data sources, thereby improving the overall performance of forecasting models. [20]

The continuous distribution assumption of Word2Vec further enhances its applicability in power systems. In electricity demand forecasting, the continuity of historical data is crucial for understanding trends in load variations. By learning Word Embeddings from historical load data, the model can better capture these trends, ultimately improving the accuracy of predictions regarding future demand changes.

2.2 Transformer Model for Time Series Data Forecasting

The Transformer model is an advanced neural network structure designed for time series data prediction, initially crafted for natural language processing [27]. Its successful applications in fields like electricity demand forecasting highlight its broad adaptability. The model's core principles encompass self-attention mechanism, multi-head attention, positional encoding, residual connections, and layer normalization. The self-attention mechanism enables the model to allocate different attention to various parts of the input sequence, aiding in capturing long-range dependencies. Multi-head attention, through parallel processing of multiple heads, empowers the model to learn features at multiple levels and types. Positional encoding introduces position information of elements in the input sequence, aiding in handling temporally ordered time series data. Residual connections and layer normalization enhance the model's training stability. In electricity demand forecasting tasks, the Transformer model demonstrates distinct advantages. Its parallelism facilitates efficient processing of long sequences, while the capability to capture global information contributes to a more accurate understanding of overarching factors like seasonality and periodicity. Additionally, the model exhibits strong adaptability to sequence lengths, allowing flexible handling of inputs across different time scales. Its straightforward structure and scalability enable researchers to tailor the model according to task requirements, meeting specific demands in electricity demand forecasting. [28]

However, despite the evident advantages of the Transformer model in electricity demand forecasting tasks, several challenges and limitations persist. Firstly, the model's computational complexity, particularly for large-scale time series data, may result in high training and inference costs, posing challenges for real-time applications and resource-constrained environments. Secondly, the model may be sensitive to noise and outliers in the input sequence, lacking robustness to handle anomalous data. Furthermore, while the model generally handles long-term dependencies well, challenges may arise in extreme cases, leading to gradient vanishing or exploding issues. Additionally, the Transformer model may not perform well in training with small sample data, making it prone to overfitting. In electricity demand forecasting tasks, where data for specific regions or time points may be limited, this could result in suboptimal generalization performance on such data. Addressing these challenges provides avenues for further research to enhance and optimize the Transformer model for the specific requirements of electricity system applications.

2.3 Attention Mechanism for Multi-Source Data Fusion

The attention mechanism [29]is an advanced neural network structure designed for handling multi-source data fusion tasks. Its core principle involves parallel processing of input data through attention mechanism, focusing on capturing different feature relationships. In the context of data fusion for electricity demand prediction tasks, this mechanism demonstrates significant advantages. Firstly, the attention mechanism exhibits powerful capabilities in handling multi-modal data fusion. In electricity demand prediction tasks, various data sources contribute to factors such as weather, seasonality, holidays, among others, and these factors may exist in different forms. The attention mechanism effectively integrates these heterogeneous data sources, enabling the model to simultaneously focus on key information from each data source that has a significant impact on demand. This flexibility enhances its adaptability, allowing it to handle the complex multi-modal information present in power systems. Secondly, the mechanism excels in flexibility and generalization capabilities. Its ability to process multiple source data in parallel provides the model with increased flexibility to adapt to the heterogeneity among different data sources. This enhances the model's generalization ability to new data and diverse conditions, contributing to its adaptability to dynamic changes in power systems. This generalization capability is particularly crucial in electricity demand prediction, given the various influencing factors and changing conditions. Thirdly, the attention mechanism can capture feature relationships at multiple hierarchical levels. In electricity demand prediction, there exist short-term and long-term relationships among influencing factors. The multi-level capturing ability of the attention mechanism allows the model to simultaneously consider these relationships, leading to a more comprehensive understanding of the formation mechanisms of electricity demand. This contributes to improving the model's ability to model complex time series data. Lastly, the representation capacity of the model is enhanced by aggregating the outputs from multiple heads. The mechanism provides a richer and more comprehensive data representation, aiding the model in capturing essential features within the data. In electricity demand prediction, this implies that the model can more accurately comprehend global influencing factors such as seasonality and periodicity, thereby improving prediction accuracy and robustness[30].

3 Methods

3.1 Overview

The model employed in this study is an innovative deep learning architecture that integrates Word2Vec, ConvTrans, and Attention Mechanism with the aim of enhancing the accuracy of electricity demand forecasting. The fundamental process of the model encompasses data collection, preprocessing, Word2Vec embedding, historical demand data processing, ConvTrans layer, and Attention Mechanism.

Initially, a comprehensive dataset was constructed by gathering geographical information data containing various factors such as historical load data, weather information, quarters, holidays, months, geographical locations, population quantities, etc. These data underwent preprocessing to ensure alignment on the same time axis. Subsequently, the Word2Vec embedding layer was applied to transform diverse influencing factors like weather and quarters into distributed representations, capturing their semantic information. Historical load data underwent embedding through the Word2Vec model using a sliding window approach, incorporating them into the same vector space. Following that, the Transformer encoder layer was introduced, leveraging the encoder structure of the Transformer model to process all embedded sequences. This aids in capturing long-term dependencies within the sequences, enhancing the modeling effectiveness for historical load and various factors. Additionally, the model incorporated an Attention Mechanism, enabling parallel attention to complex relationships among different influencing factors. This model, through the comprehensive application of these advanced techniques, achieves more accurate predictions of electricity demand variations. By effectively utilizing multi-source data and considering the interactive relationships among various influencing factors, it provides an optimized deep learning model for intelligent forecasting of energy demand in the power system. The structure of the model is shown in Fig.1.

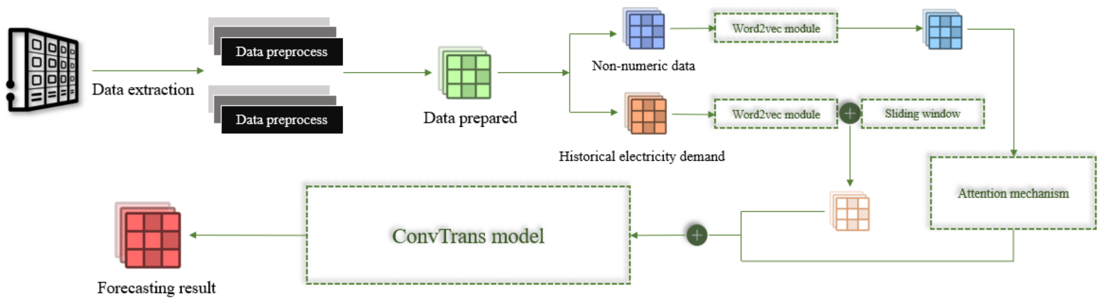


Fig.1 Model Structure Diagram.

3.2 Data Collection

To conduct electricity demand forecasting, we executed a thorough data collection process to ensure that the model's training and predictions are based on a comprehensive information foundation. The following are the specific steps and detailed procedures involved in the data collection: (1) Data Sources: Initially, multiple data sources were identified in Beijing's Chaoyang district, including the Beijing Chaoyang District Power Company, meteorological bureau, statistical bureau, and other relevant government agencies that voluntarily disclose data. These sources provided data on the Chaoyang district's electricity system, weather, and socio-economic aspects. (2) Historical Load Data: Historical load data for the past five years were obtained from the Chaoyang District Power Company. This dataset includes detailed hourly electricity consumption, providing a time-series representation of the power system load. (3) Weather Data: Meteorological data for the same time range were obtained from voluntary disclosures by the Chaoyang district meteorological bureau. This data encompassed temperature, humidity, wind speed, among other factors, and was utilized to analyze the impact of weather factors on electricity demand. (4) Quarter, Month, and Holiday Information

Integration: Information regarding quarters, months, and specific holidays was integrated into the dataset. (5) Geographical Information and Population Data: For this data collection, we used data from the Chaoyang District Statistical Bureau, including area divisions, population figures, and electricity usage characteristics.

Strict data cleaning procedures were implemented, addressing missing values, outliers, and anomalies. Interpolation methods were applied to fill missing data, ensuring data completeness. Subsequently, all data were aligned and synchronized on the time axis to ensure temporal consistency across different data sources. For data storage and management, an SQL-based database system was adopted, organizing all collected data into tables that store information on load, weather, and other influencing factors. Throughout the entire process, comprehensive data quality checks were conducted to ensure accuracy and consistency. An additional layer of validation and processing was applied to handle any identified abnormal data. This meticulous data collection process provides a robust foundation for developing an accurate electricity demand forecasting model for the Chaoyang district. The electricity demand data collection process is shown in Fig.2.

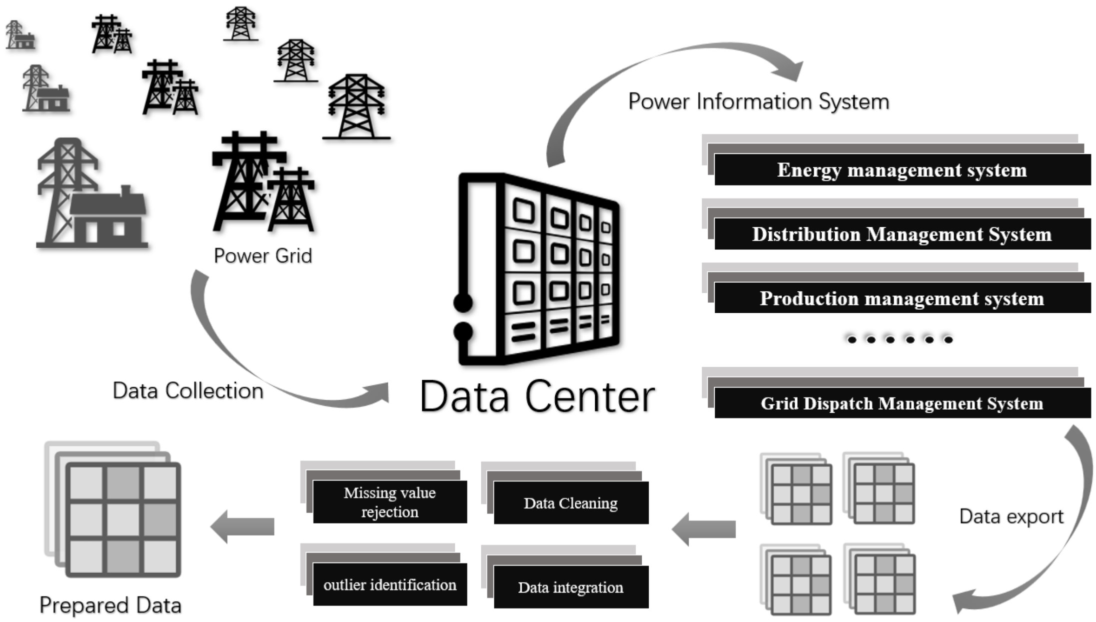


Fig.2 Electricity load data collection process.

3.3 Word2vec and Sliding Windows for Power System Data Preprocessing

The crucial innovative step in the proposed model lies in the application of a sliding window approach to handle the time series data of electricity demand. The entire time series is divided into fixed-size windows, each covering a continuous segment of the time series data, forming a subsequence. These windows slide along the time axis, covering the entire dataset. The flexibility of this step allows us to capture the characteristics of electricity load variations at different time scales. Subsequently, for each subsequence generated by the sliding window, we employed the Word2Vec model for embedding. Historical electricity load data is treated as a sequence of words inputted into the Word2Vec model. The model, by learning contextual information, maps this sequence into Word Embeddings within a high-dimensional vector space. Essentially, this process transforms historical load data into vector representations imbued with semantic information. Finally, we assemble the Word Embeddings generated for each sliding window in chronological order to create a Word Embedding time series. This time series preserves the semantic evolution of historical electricity load data while considering the continuity of time. Through this integrated processing pipeline of Word2Vec and sliding window techniques, we establish a more informative and semantically rich data representation, providing the electricity demand forecasting model with more accurate and comprehensive inputs. The principle of word2vec and sliding windows for power system data preprocessing is shown in Fig.3.

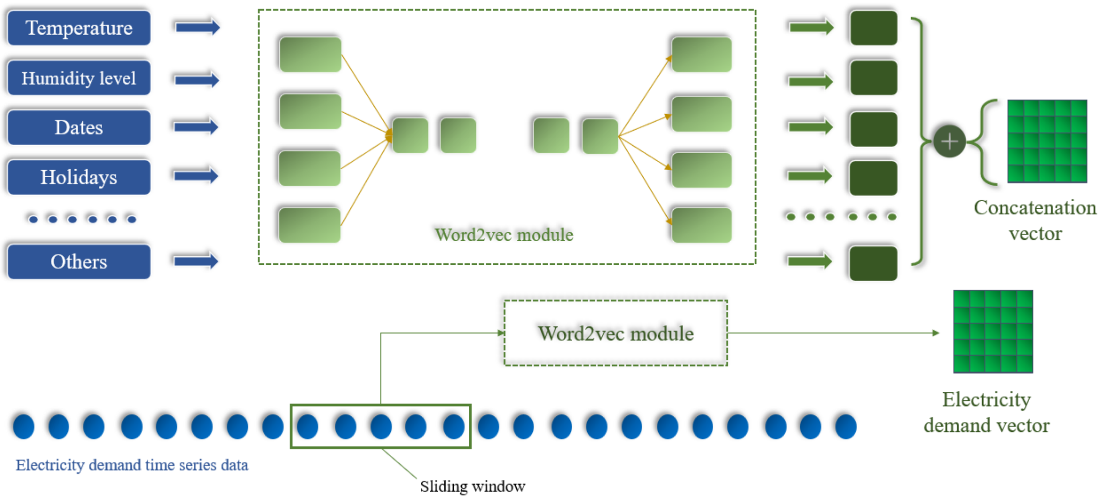


Fig.3 The principle of word2vec and sliding windows for power system data preprocessing.

3.4 Multiple Attention Mechanisms for Data Fusion of Multiple Influences

Due to the significant correlation between power demand factors and time, similar influencing factors may have entirely different implications on different dates. Leveraging this characteristic, this paper introduces an attention module to adaptively extract date-related information from influencing factors, facilitating the fusion of feature representations between influencing factors and dates.

Initially, the attention module takes an influencing factor vector obtained by concatenating multiple factor values and a date vector as input. Three linear layers are employed to obtain variables Q, K, and V. The 32-dimensional date vector is fed into the first linear layer to yield a 16-dimensional query vector, denoted as Q. Simultaneously, the 5-dimensional influencing factor vector is input into the other two linear layers to obtain 16-dimensional vectors, designated as keys (K) and values (V). After obtaining these three vectors, Q and K are utilized to calculate attention weights, as per the formula (1).

$$Attention = \mathrm{Softmax}(Q^T \cdot K) \qquad \text{[Formular 1]}$$

The transposed Q is matrix-multiplied with K, resulting in a matrix of shape (16, 16). Applying the Softmax function to the last dimension of the matrix yields attention weights, denoted as Attention. In this module, Q represents features extracted from the date vector, while K captures features extracted from the influencing factor vector. The calculated attention weights, Attention, thus integrate information from both influencing factors and dates. Finally, as expressed in formula (2), the product of V and the transposed attention weights yields the output of the attention module.

$$\mathrm{Output} = \mathrm{V} \cdot \mathrm{Attention}^{\mathrm{T}} \qquad \text{[Formular 2]}$$

In this step, attention weights are employed to adjust the weights of various feature components within the influencing factor information. Consequently, the attention module effectively utilizes the correlation between dates and influencing factors to extract pertinent hidden features of influencing factors. The principle of Multiple attention mechanisms for data fusion of multiple influence data is shown in Fig.4.

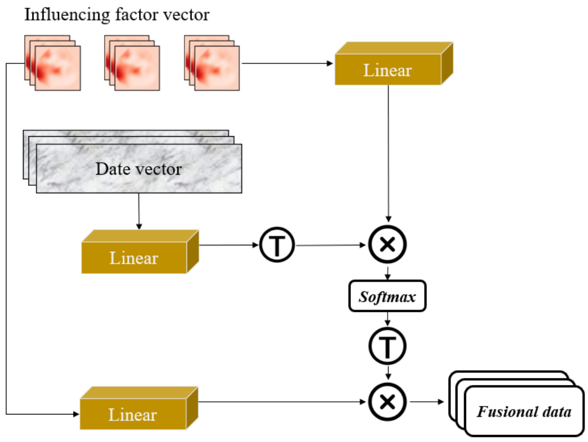


Fig.4 The principle of Multiple attention mechanisms for data fusion of multiple influence data.

3.5 Convtrans Model for Electricity Demand Forecasting

The Transformer model is a deep learning architecture based on attention mechanisms, originally proposed by Vaswani et al. in 2017 for natural language processing tasks such as machine translation. The ConvTrans model is an improved model based on the Transformer model, specifically for time series data forecasting. Its core principles of the Transformer model include the following four aspects.

(1) Self-Attention Mechanism: The Transformer model's core is the self-attention mechanism. In traditional Recurrent Neural Networks (RNNs), information is sequentially passed through the sequence, while the self-attention mechanism allows the model to assign varying attention to different positions in the input sequence. This enables the model to better capture long-range dependencies.

(2) Multi-Head Attention Mechanism: The Transformer introduces the multi-head attention mechanism, processing multiple heads in parallel. This allows the model to learn features at multiple levels and types. Each head focuses on different aspects of the sequence, enhancing the model's expressive power.

(3) Positional Encoding: Since the Transformer lacks inherent sequential information like traditional RNNs, positional encoding is introduced to embed the position information of elements in the input sequence into the model. This is crucial for handling time-ordered time series data.

(4) Residual Connection and Layer Normalization: Each sub-layer in the model includes residual connections and layer normalization, contributing to improved training stability. Residual connections alleviate the vanishing gradient problem by directly adding the input to the output of the sub-layer.

The ConvTrans model improves the calculation of Attention based on the Transformer to adapt to time series data and proposes the Convolutional Self-Attention algorithm to address the Transformer's poor scalability issue. The core of the Convolutional Self-Attention mechanism is Self-Attention, which maps the input feature map of the current time step to two different feature spaces, calculates query, key, and value, computes similarity scores, and finally obtains summarized features through normalization and weighted summation. The Convolutional Self-Attention mechanism introduces convolutional operations, allowing the model to better capture local context information in time series, thereby reducing the impact of anomalies on prediction results and enhancing the modeling capability of the model. Additionally, the ConvTrans model possesses unique advantages of the Transformer architecture, such as supporting parallelization, faster training, and stronger long-term dependency modeling capabilities, resulting in improved performance on long sequences. The principle of ConvTrans model for electricity demand forecasting is shown in Fig.5.

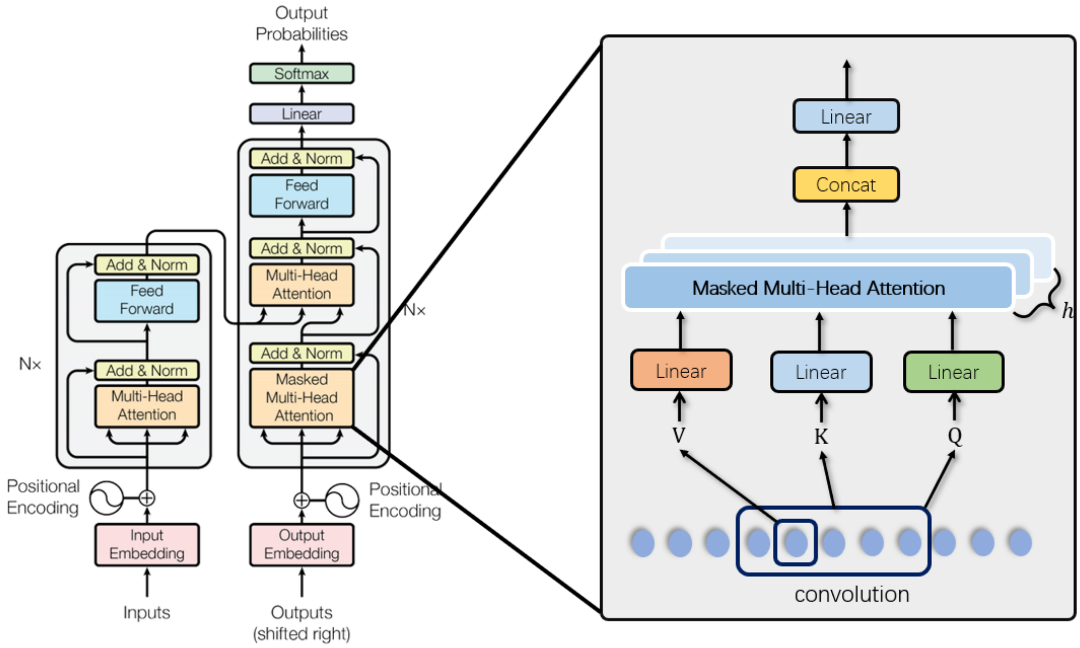


Fig.5 The principle of ConvTrans model for electricity demand forecasting.

The logic of the entire model is represented in pseudo-code as follows:

Algorithm 1: *Integrating Word2Vec, ConvTrans, and Attention Mechanisms for Electricity Demand Forecasting*

```
# input layers
input_date = layers. Input(shape=(date_vector_size,))
input_factors = layers.Input(shape=(factors_vector_size,))

# Embedding layer by using Word2Vec
embedding_layer = layers.Embedding(input_dim, input_factors)

# Attention module
Q = layers.Dense(16)(input_date)
K = layers.Dense(16)(embedding_layer)
V = layers.Dense(16)(embedding_layer)
attention_weights = softmax(matmul(transpose(Q), K) / sqrt(16))
attention_output = matmul(attention_weights, V)
# Concatenate attention output with original date vector
concatenated_output = Concatenate(attention_output, input_date)

# ConvTrans layers for electricity demand forecasting
forecaste_output=ConvTrans(input_date, concatenated_output)

# Output layer
output = layers.Dense(output_size, activation='linear', forecaste_output)

# Build the model
model =Model(inputs, outputs)

# Compile the model
model. Train()
```

4 Experience

4.1 Experimental Design

This study conducted three experiments in total. The first experiment involved contrasting our proposed method with various baseline methods on a single dataset, confirming the superior performance of our approach. The second experiment compared the operational results of our method across multiple datasets, demonstrating the robustness and generalization capability of our approach. Additionally, a model ablation experiment was conducted for our proposed method, affirming the intrinsic value of each component in the model.

The experiments were conducted on a workstation equipped with an Intel Core i7 processor and 32GB of memory. The software environment included Python 3.8, TensorFlow 2.5, and relevant deep learning libraries. GPU acceleration, facilitated by an NVIDIA GeForce RTX 3080 graphics card, was utilized for model training. All experiments were conducted on the Ubuntu 20.04 LTS operating system.

For our model, meticulous parameter tuning was performed to ensure experiment comparability and result robustness. The model parameters were set as follows:

1. Word vector dimension: 300.s
2. sliding window size: 5.
3. Learning rate: 0.001.
4. epoch: 50.
5. Batch size: 64.

4.2 Experimental Data Set

In the multi-dataset comparative experiment, we selected four public datasets, namely, the Global Energy Forecasting Competition (GEFCom), REDI: Spanish Electricity Demand dataset, PJM Hourly Energy Consumption Data, and EIA-930, for evaluation:

DataSet #1. Global Energy Forecasting Competition (GEFCom) [31]: GEFCom is a collection of datasets provided by the International Energy Forecasting Competition, primarily covering electricity load forecasting and renewable energy generation prediction. These datasets include electricity demand and production data from different geographic regions and time periods, featuring high resolution and diversity. The GEFCom dataset covers characteristics of power systems globally, enhancing the generalization capability of the model proposed in this study.

DataSet #2. REDI: Spanish Electricity Demand dataset [32]: The Spanish Electricity Demand dataset (REDI) covers electricity demand across various regions in Spain, offering detailed time series data at an hourly granularity. This dataset meticulously captures load variations in the Spanish electricity system. REDI's temporal span and regional coverage contribute to enhancing the robustness and generalization of the model in practical applications.

DataSet #3. PJM Hourly Energy Consumption Data [33]: PJM Hourly Energy Consumption Data, provided by PJM Interconnection, includes electricity demand data for most of the northeastern United States at an hourly resolution. This dataset provides information on power demand over an extended time span. The PJM dataset encompasses a crucial region in the U.S., facilitating the study of seasonal patterns, trends, and special events in power demand, providing rich learning material for the model.

DataSet #4. EIA-930 [34]: EIA-930, offered by the U.S. Energy Information Administration, consists of monthly electricity demand data for various regions in the United States, offering nationwide and regional-level demand information. This dataset provides comprehensive and detailed information for studying U.S. electricity demand trends and patterns, as well as predicting future demand changes.

These four datasets provide diverse data sources for the electricity demand forecasting task, covering different regions and time periods globally. Leveraging these datasets enables the establishment of robust and accurate electricity demand prediction models. It allows a deep understanding of the operational conditions of power systems under various circumstances, enhancing the model's generalization performance to adapt to different geographic and environmental features. In this study, the GEFCom dataset was utilized for the performance comparison experiment on a single dataset.

For preprocessing the time series data on electricity demand, common techniques such as data normalization, standardization, missing value imputation, and anomaly detection were employed. The dataset was then partitioned into training, validation, and test sets in a 7:2:1 ratio. In the ablation experiment, we systematically removed certain key components from the model to assess their impact on model performance.

4.3 Baseline Model and Evaluation Indicators

In this study, six SOTA models were found from the literature as baseline models. The models are as follows:

Model #1: [35] this study proposed a novel optimal hybrid strategy for building load prediction that combines BiLSTM, CNN, and grey wolf optimization (GWO).

Model #2: [36] this study used the Multiple Seasonal-Trend Decomposition using Loess (MSTL) technique for the electricity demand forecasting problem.

Model #3: [37] this study presented a model that integrates a multi-criteria approach which provides the selection of relevant independent variables and artificial neural networks to forecast the electricity demand in countries.

Model #4: [38] this study proposed a hybrid combination technique, based on a deep learning model of Convolutional Neural Networks and Echo State Networks, named as CESN

Model #5: [39] this study provided a solution based on statistical methods (ARIMA, ETS, and Prophet) to predict monthly power demand, which approximates the relationship between historical and future demand patterns.

Model #6: [40] this study proposed an ideal technique for short-term power demand prediction as a novel hybrid approach comprising two distinct methods, namely the Elman neural network (ELM) and adaptive network–based fuzzy inference system (ANFIS).

To compare model performance differences, the following model comparison indicators were used:

1. Root Mean Squared Error (RMSE): Measures the average difference between actual and predicted values, with lower values indicating better performance.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \text{[Formular 3]}$$

where n is the sample size, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

2. Mean Absolute Error (MAE): Represents the average absolute difference between actual and predicted values, with smaller values indicating better performance.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad \text{[Formular 4]}$$

3. Mean Squared Error (MSE): The square of RMSE, used to quantify the squared average difference between actual and predicted values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \text{[Formular 5]}$$

4. Coefficient of Determination (R^2): Represents the percentage of variance explained by the model, with values closer to 1 indicating better explanatory power.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{[Formular 6]}$$

where $\bar{y}$ is the mean of actual values.

5. Pearson Correlation Coefficient: Measures the linear correlation between actual and predicted values, ranging from -1 to 1, with values closer to 1 indicating stronger correlation.

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \text{[Formular 7]}$$

where $\mathrm{Cov}(X, Y)$ is the covariance between X and Y, and σ_X and σ_Y are the standard deviations of X and Y.

6. Accuracy: Represents the proportion of correctly predicted samples out of the total samples, commonly used in classification tasks.

$$\mathrm{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{[Formular 8]}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

7. Precision: Represents the proportion of correctly predicted positive instances out of all predicted positive instances, particularly useful for imbalanced datasets.

$$\mathrm{Precision} = \frac{TP}{TP+FP} \quad \text{[Formular 9]}$$

8. Recall: Represents the proportion of actual positive instances correctly predicted by the model out of all actual positive instances.

$$\mathrm{Recall} = \frac{TP}{TP+FN} \quad \text{[Formular 10]}$$

9. F1-score: A metric that combines precision and recall, suitable for imbalanced class situations.

$$\text{F1-score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad \text{[Formular 11]}$$

10. Mean Absolute Percentage Error (MAPE): Measures the average percentage difference between actual and predicted values, a commonly used percentage error metric.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad \text{[Formular 12]}$$

where n is the sample size, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

4.4 Comparison Experiment

4.4.1 Comparison results with SOTA models

Our model demonstrates a notable performance advantage in the electricity demand prediction comparative experiment. The results of the comparison with the six baseline models are shown in Table 1.

Table 1. The results of the comparison with the six baseline models.

Model	MSE	RMSE	MAE	MAPE	R^2	Pearson Correlation	F1-score
Model #1[35]	9.75%	320.03%	7.02%	8.50%	77.54%	80.98%	77.21%
Model #2[36]	8.72%	445.39%	8.89%	7.25%	83.19%	84.09%	79.29%
Model #3[37]	9.66%	470.85%	8.53%	**6.86%**	88.93%	88.57%	83.11%
Model #4[38]	7.14%	413.38%	8.41%	6.70%	88.39%	76.76%	**89.99%**
Model #5[39]	7.74%	405.97%	7.17%	9.03%	77.81%	87.64%	81.20%
Model #6[40]	**5.09%**	**488.59%**	6.40%	8.19%	88.43%	76.73%	78.20%
Ours	5.81%	463.56%	**5.91%**	6.83%	**89.24%**	**91.62%**	**90.12%**

Source: By authors.

Leveraging an attention module designed to adaptively extract date-related information from influencing factors, our model effectively fuses features from both influencing factors and dates, enhancing the model's ability to capture nuanced dependencies in the data. The attention mechanism enables the adjustment of the weightings of different features within influencing factors, effectively incorporating the temporal correlation of electricity demand. This robust fusion of information contributes to the superior generalization and performance of our proposed model compared to the alternative methodologies, highlighting its efficacy in addressing the intricacies of electricity demand forecasting. Figure 6 illustrates the indicators that measure the accuracy of the models' forecast.

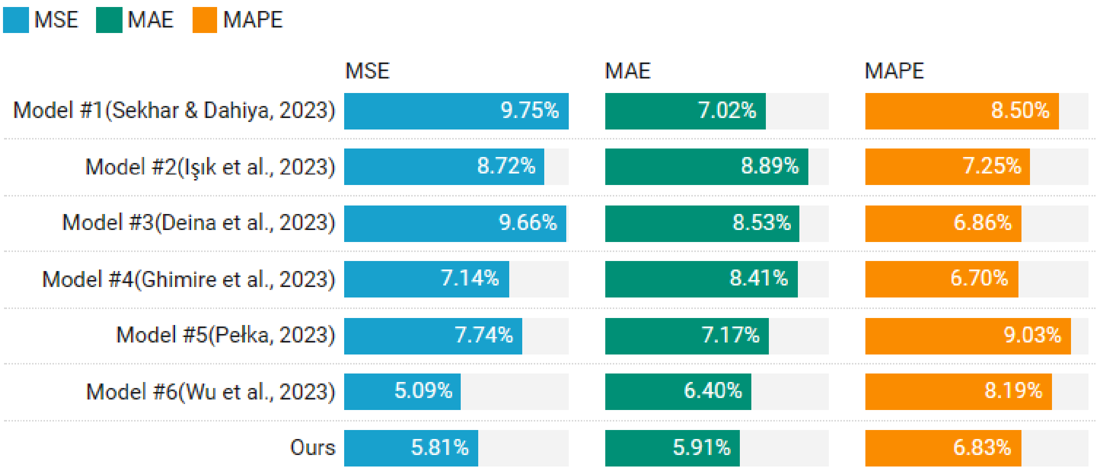


Fig.6 Comparison of model accuracy indicators.

Figure 7 illustrates the indicators that measure the comprehensive performance of the models' forecast.

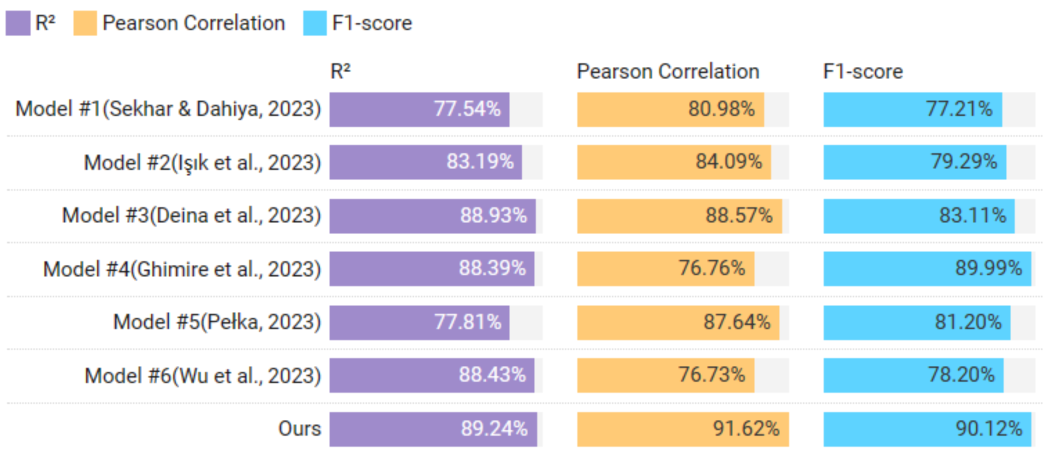


Fig.7 Comparison of model composite indicators.

4.4.2 Comparison results in different datasets

To validate the robustness and generalization capabilities of the proposed model, we conducted extensive and in-depth performance tests across the four distinct electricity demand datasets and the dataset collected by ourselves. The results of this series of experiments clearly demonstrate that the model performs exceptionally well in diverse scenarios, showcasing outstanding robustness and generalization abilities. Table2 shows the comparison results in four datasets.

Table 2. The results of the comparison within different datasets.

Data sets	MSE	RMSE	MAE	MAPE	R^2	Pearson Correlation	F1-score
GEFCom	5.81%	463.56%	5.91%	6.83%	**89.24%**	**91.62%**	90.12%
REDI	5.51%	**481.35%**	5.24%	**9.15%**	73.51%	82.15%	**90.15%**
PJM	6.52%	421.03%	**6.41%**	5.67%	72.31%	86.62%	87.35%
EIA-930	**4.25%**	419.12%	5.12%	6.23%	79.26%	81.12%	82.51%
Real data	5.35%	425.72%	5.51%	6.72%	82.52%	81.15%	83.46%

Source: By authors.

Initially, we conducted tests on the Global Energy Forecasting Competition (GEFCom) dataset. Through performance testing on this diverse dataset, we were able to assess the model's generality in predicting electricity demand under different geographical and temporal conditions. The experimental results indicate that the proposed model excelled on the GEFCom dataset, affirming its strong generalization performance.

Subsequently, we performed further validation on the REDI: Spanish Electricity Demand dataset. By conducting performance tests on this regional dataset, we evaluated the model's adaptability to different geographical regions. The results showed satisfactory performance on the REDI dataset, indicating robustness in predicting electricity demand across diverse geographic areas. Figure 8 shows the results of the model run on this dataset and the differences on the GEFCom dataset.

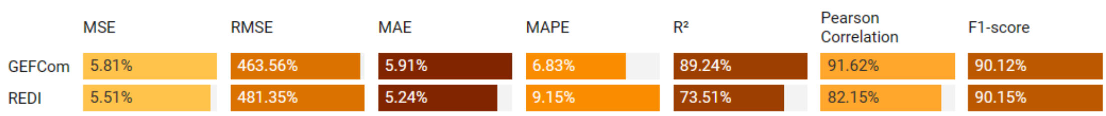


Fig.8 Result difference between REDI and GEFCom dataset.

Next, we turned to the PJM Hourly Energy Consumption Data, conducting tests on this pivotal dataset to examine the model's predictive capabilities for specific regional electricity demand. The experimental results demonstrated excellent performance on the PJM dataset, highlighting the model's ability to adapt to distinct geographical features. Figure 9 shows the results of the model run on this dataset and the differences on the GEFCom dataset.

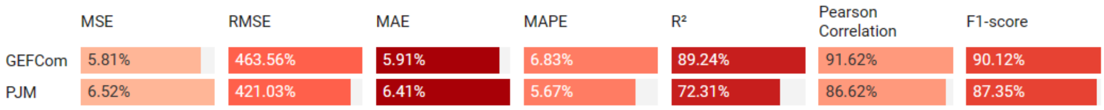


Fig.9 Result difference between PJM and GEFCom dataset.

Then, we utilized the EIA-930 dataset provided by the U.S. Energy Information Administration, conducting tests on this nationwide dataset to assess the model's adaptability to larger-scale data. The results indicated that the proposed model maintained robust performance on the EIA-930 dataset, showcasing its superior generalization capabilities in larger-scale datasets. Figure 10 shows the results of the model run on this dataset and the differences on the GEFCom dataset.

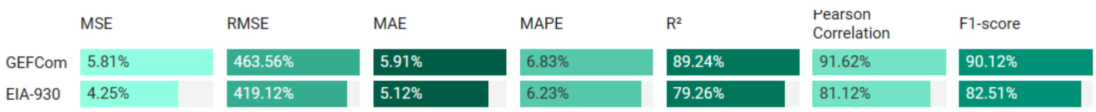


Fig.10 Result difference between EIA-930 and GEFCom dataset.

Finally, we utilized the collected dataset by ourselves. It is the electricity demand data for Chaoyang District, Beijing, China, November 25, 2023. Conducting tests on this specific dataset to assess the model's adaptability to real data. The results indicated that the proposed model maintained robust performance on the real dataset, showcasing its superior generalization capabilities in real datasets. Figure 11 shows the results of the model run on this real dataset and the differences on the GEFCom dataset.

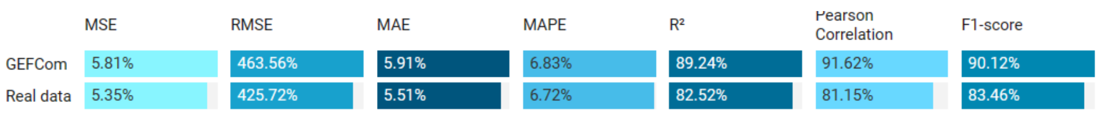


Fig.11 Result difference between real dataset and GEFCom dataset

Synthesizing the results from these four datasets, our conclusion is that the proposed model exhibits exceptional performance across different geographical regions and temporal conditions, demonstrating remarkable robustness and generalization capabilities.

4.4.3 Comparative results of Flops on different data sets

We also compared the Flops values of our proposed model and six baseline models during the inference process. The results indicate that our model has lower Flops values during inference, demonstrating the model's reduced computational overhead. The comparison results are shown in Figure 12.

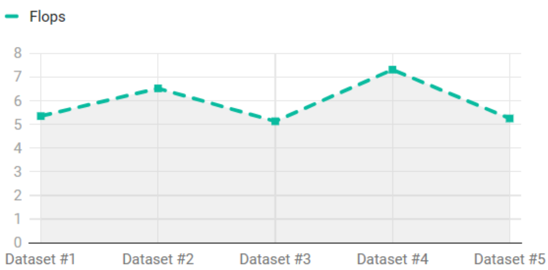


Fig.12 The Flops(h) comparison results.

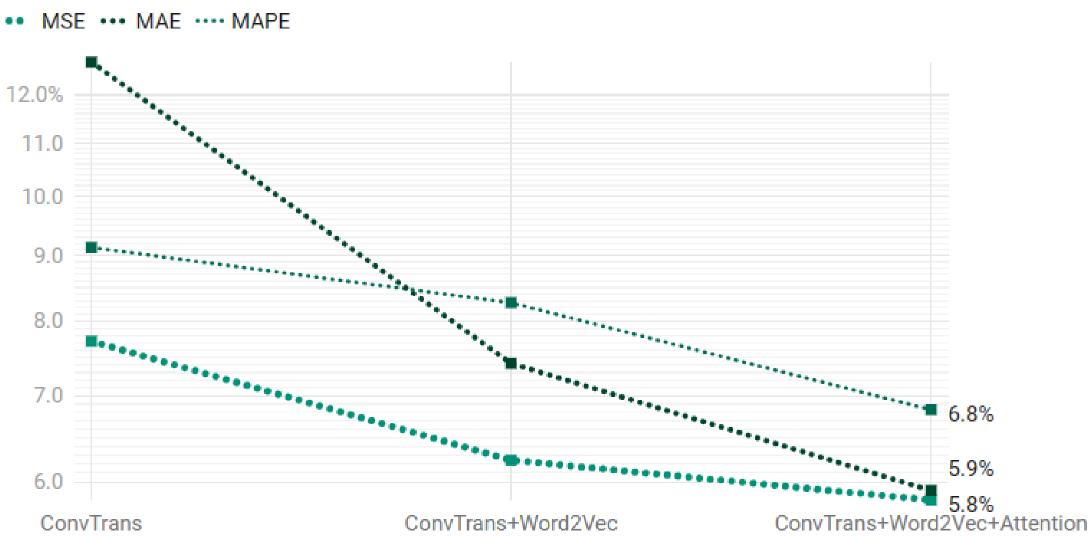
4.5 Ablation Experiment

To validate the effects produced by the individual models in the proposed model, an ablation experiment is concluded that rejects each module separately and measures the change in model performance. The results of the experiment are shown in Table 3.

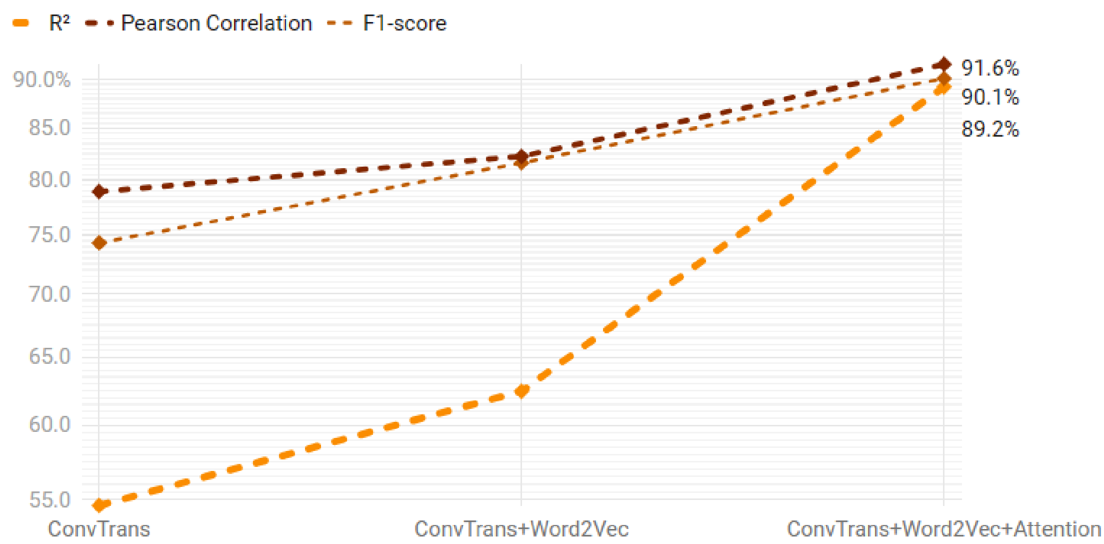
Table 3. Results of the ablation experiment.

ConvTrans	Word2Vec	Attention	MSE	RMSE	MAE	MAPE	R^2	Pearson Correlation	F1-score
✓			7.72%	376.93%	12.72%	9.13%	54.62%	78.92%	74.31%
✓	✓		6.24%	362.32%	7.42%	8.27%	62.46%	82.25%	81.63%
✓	✓	✓	5.81%	463.56%	5.91%	6.83%	89.24%	91.62%	90.12%

The results in Table 3 are presented graphically as Figure 13.



(a) comparison result of MSE, MAE, MAPE indicators.



(b) comparison result of R^2 , Pearson Correlation, F1-score indicators.
Fig.13 Ablation experiment results.

We conducted experiments by progressively excluding the attention module to compare the performance variations of the model when the attention mechanism is eliminated. The results indicate that the introduction of the attention module significantly enhances the model's predictive accuracy. The ablation experiments further validate the critical role of this module in the model, as it can adaptively integrate date and external influencing factor information, thereby aiding the model in capturing the spatiotemporal relationships of electricity demand more effectively. In the absence of the attention module, the assistance provided by external influencing factor data in enhancing the model's predictive capabilities is noticeably reduced, leading to a decrease in predictive performance.

We also conducted ablation experiments by progressively removing the Word2Vec module to compare the performance variations of the model after excluding the Word2Vec module. The results indicate that the introduction of the Word2Vec module significantly enhances the model's semantic representation capabilities and feature learning effectiveness. The ablation experiments further validate the critical role of this module in the model. We compared the performance of the model in semantic representation and feature learning when the Word2Vec module was removed. The results show that the elimination of the Word2Vec module noticeably restricts the model's expression of semantic information for vocabulary, leading to a decrease in model performance. The presence of the Word2Vec module enables the model to better understand the relationships between words, thereby improving its performance in the task of electricity demand forecasting.

5 Conclusion

This study aims to enhance the accuracy and generalization capability of electricity demand forecasting, thereby improving the efficiency of electricity resource production, distribution, and utilization. Through innovative designs such as the introduction of attention and Word2Vec modules, we have constructed a high-performance electricity demand prediction model. Extensive experimental validations and testing on multiple public datasets demonstrate the exceptional performance of our model in electricity demand prediction tasks.

The incorporation of the attention module enables the model to dynamically fuse date and external factors, better capturing the spatiotemporal relationships between electricity demand and external influences. Additionally, the application of the Word2Vec module enhances the understanding and representation of semantic information, allowing the model to capture complex data features more accurately. The use of the ConvTrans model further strengthens the modeling capability for time series data, increasing sensitivity to variations in electricity demand time series.

The combination of these innovative designs positions our model to outperform six baseline models across various performance metrics. Comparative experiments against the baseline models on multiple public datasets, including the Global Energy Forecasting Competition, REDI, PJM Hourly Energy Consumption Data, and EIA-930, reveal outstanding robustness and generalization capabilities. The model exhibits superior performance in metrics such as Mean Absolute Error (MAE), Coefficient of Determination (R^2), Pearson Correlation Coefficient, and F1-score compared to the baseline models. During the inference process, the model demonstrates lower Flops values, indicating relatively lower computational overhead.

In summary, our proposed electricity demand prediction model not only showcases innovation in design through the incorporation of attention mechanisms, Word2Vec modules, and ConvTrans models but also significantly improves performance metrics. The model's outstanding robustness and generalization capabilities across multiple datasets provide reliable decision support for electricity system planning and operations.

6. Outlook

Although this study has achieved significant advancements in the field of electricity demand forecasting, there are still notable limitations that should be acknowledged. Firstly, the performance of the model is constrained by the diversity and coverage of the adopted datasets. Despite extensive testing across multiple datasets, these may not comprehensively cover all potential scenarios in electricity demand. Consequently, the generalization ability of the model in specific geographic regions or under environmental conditions requires further in-depth validation. To address these limitations and advance research in electricity demand forecasting, future studies can explore the incorporation of additional external factors from diverse sources and dimensions, such as socioeconomic factors and policy changes. This extension aims to enhance the model's capability to model complex influencing factors comprehensively, thereby improving its accuracy in predicting electricity demand variations.

Secondly, the sensitivity of the model to external factors may be limited, especially when facing extreme weather events or other unforeseen circumstances. For these complex situations, the model may benefit from more sophisticated approaches involving the introduction of complex external factors or specialized handling strategies to enhance its capability in responding to unforeseen events. Consequently, future research could focus on optimizing the model's real-time performance and adaptability, particularly in dealing with unexpected events. This could involve the incorporation of dynamic update mechanisms and more flexible model structures to ensure efficient and accurate predictions in rapidly changing electricity system environments.

Lastly, customization of the model to meet specific regional or industry requirements can be explored further. This tailoring process would adapt the model to the unique characteristics of different backgrounds in electricity demand forecasting tasks. Through these customized adjustments, the model could achieve higher relevance and applicability in diverse contexts.

Exploring these future research directions is anticipated to expand the applicability of electricity demand forecasting models, enhancing their accuracy and operability. This, in turn, would provide robust support for the efficient production, distribution, and consumption of electric energy.

Reference

[1] Saglam, M., Spataru, C. and Karaman, O.A. Electricity demand forecasting with use of artificial intelligence: The case of Gokceada Island. Energies, 2022, 15(16), 5950.

[2] Deng, X. and Lv, T. Power system planning with increasing variable renewable energy: A review of optimization models. Journal of Cleaner Production, 2020, 246, 118962.

[3] Salman, R.A., Myeongbae, L., Jonghyun, L., Cho, Y. and Changsun, S. A comparative study of energy big data analysis for product management in a smart factory. Journal of Organizational and End User Computing, 2022, 34(2), 1-17.

[4] Tawn, R. and Browell, J. A review of very short-term wind and solar power forecasting. Renewable and Sustainable Energy Reviews, 2022, 153, 111758.

[5] Nia, A.R., Awasthi, A. and Bhuiyan, N. Industry 4.0 and demand forecasting of the energy supply chain: A literature review. Computers and Industrial Engineering, 2021, 154, 107128.

[6] Philip, K., Jammeh, K. and Oduor, J. An outlook of energy demand, supply and cost in Tanzania. 2022.

[7] Li, J., Ho, M.S., Xie, C. and Stern, N. China's flexibility challenge in achieving carbon neutrality by 2060. Renewable and Sustainable Energy Reviews, 2022, 158, 112112.

[8] Nam, K., Hwangbo, S. and Yoo, C. A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea. Renewable and Sustainable Energy Reviews, 2020, 122, 109725.

[9] Somu, N., MR, G.R. and Ramamritham, K. A deep learning framework for building energy consumption forecast. Renewable and Sustainable Energy Reviews, 2021, 137, 110591.

[10] Aslam, S., Herodotou, H., Mohsin, S.M., Javaid, N., Ashraf, N. and Aslam, S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. Renewable and Sustainable Energy Reviews, 2021, 144, 110992.

[11] Elharrouss, O., Akbari, Y., Almaadeed, N. and Al-Maadeed, S. Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. arXiv preprint, 2022, arXiv:2206.08016.

[12] Kothai Andal, C. and Jayapal, R. Intelligent power and cost management system in small-scale grid-associated PV-wind energy system using RNN-LSTM. SN Computer Science, 2023, 4(6), 746.

[13] Zha, W., Liu, Y., Wan, Y., Luo, R., Li, D., Yang, S. and Xu, Y. Forecasting monthly gas field production based on the CNN-LSTM model. Energy, 2022, , 124889.

[14] Yu, J., de Antonio, A. and Villalba-Mora, E. Deep learning (CNN, RNN) applications for smart homes: a systematic review. Computers, 2022, 11(2), 26.

[15] Algiriyage, N., Prasanna, R., Stock, K., Doyle, E.E. and Johnston, D. Multi-source multimodal data and deep learning for disaster response: a systematic review. SN Computer Science, 2022, 3, 1-29.

[16] Soydaner, D. Attention mechanism in neural networks: where it comes and where it goes. Neural Computing and Applications, 2022, 34(16), 13371-13385.

[17] Kontopoulou, V.I., Panagopoulos, A.D., Kakkos, I. and Matsopoulos, G.K. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet, 2023, 15(8), 255.

[18] Liang, F., Zhang, H. and Fang, Y. The analysis of global RMB exchange rate forecasting and risk early warning using ARIMA and CNN model. Journal of Organizational and End User Computing, 2022, 34(8), 1-25.

[19] Lima, M.A.F., Fernández Ramírez, L.M., Carvalho, P.C., Batista, J.G. and Freitas, D.M. A comparison between deep learning and support vector regression techniques applied to solar forecast in Spain. Journal of Solar Energy Engineering, 2022, 144(1), 010802.

[20] Hu, J. and Szymczak, S. A review on longitudinal data analysis with random forest. Briefings in Bioinformatics, 2023, 24(2), bbad002.

[21] Ying, C., Wang, W., Yu, J., Li, Q., Yu, D. and Liu, J. Deep learning for renewable energy forecasting: A taxonomy and systematic literature review. Journal of Cleaner Production, 2022, , 135414.

[22] Wang, S., Ren, P., Takyi-Aninakwa, P., Jin, S. and Fernandez, C. A critical review of improved deep convolutional neural network for multi-timescale state prediction of lithium-ion batteries. Energies, 2022, 15(14), 5053.

[23] Johnson, S.J., Murty, M.R. and Navakanth, I. A detailed review on word embedding techniques with emphasis on word2vec. Multimedia Tools and Applications, 2023, , 1-29.

[24] Zeng, Z., Cui, L., Qian, M., Zhang, Z. and Wei, K. A survey on sliding window sketch for network measurement. Computer Networks, 2023, 226, 109696.

[25] Ahmed, S., Nielsen, I.E., Tripathi, A., Siddiqui, S., Ramachandran, R.P. and Rasool, G. Transformers in time-series analysis: A tutorial. Circuits, Systems, and Signal Processing, 2023, 42(12), 7433-7466.

[26] Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X. and Yan, X. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Advances in Neural Information Processing Systems, 2019, 32.

[27] Xu, X.X., Wang, W., Pan, S.H. and G., S. Chinese caption of fine-grained images based on transformer. Journal of Jilin University Science Edition, 2022, 60(5), 1103-1112.

[28] Su, L., Zuo, X., Li, R., Wang, X., Zhao, H. and Huang, B. A systematic review for transformer-based long-term series forecasting. arXiv preprint, 2023, arXiv:2310.20218.

[29] Ji, B. and Zhang, Y. Few-shot relation extraction model based on attention mechanism induction network. Journal of Jilin University (Science Edition), 2023, 61(4).

[30] Wang, Q., Wang, Y., Zhang, K., Liu, Y., Qiang, W. and Han Wen, Q. Artificial intelligent power forecasting for wind farm based on multi-source data fusion. Processes, 2023, 11(5), 1429.

[31] Dolinský, J., Starovská, M. and Tóth, R. Automatic model building in GEFCom 2017 qualifying match. arXiv preprint, 2019, arXiv:1904.12608.

[32] Pegalajar, M., Ruíz, L.G.B., Cuéllar, M.P. and Rueda, R. Analysis and enhanced prediction of the Spanish electricity network through big data and machine learning techniques. International Journal of Approximate Reasoning, 2021, 133, 48-59.

[33] Sahay, K.B. and Tripathi, M.M. Day ahead hourly load forecast of PJM electricity market and ISO New England market by using artificial neural network. ISGT, 2014.

[34] Miller, G.J., Pease, G., Shi, W. and Jenn, A. Evaluating the hourly emissions intensity of the US electricity system. Environmental Research Letters, 2023, 18(4), 044020.

[35] Sekhar, C. and Dahiya, R. Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand. Energy, 2023, 268, 126660.

[36] Işık, G., Öğüt, H. and Mutlu, M. Deep learning based electricity demand forecasting to minimize the cost of energy imbalance: A real case application with some Fortune 500 companies in Türkiye. Engineering Applications of Artificial Intelligence, 2023, 118, 105664.

[37] Deina, C., dos Santos, J.L.F., Biuk, L.H., Lizot, M., Converti, A., Siqueira, H.V. and Trojan, F. Forecasting electricity demand by neural networks and definition of inputs by multi-criteria analysis. Energies, 2023, 16(4), 1712.

[38] Ghimire, S., Nguyen-Huy, T., AL-Musaylh, M.S., Deo, R.C., Casillas-Pérez, D. and Salcedo-Sanz, S. A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction. Energy, 2023, 275, 127430.

[39] Pełka, P. Analysis and forecasting of monthly electricity demand time series using pattern-based statistical methods. Energies, 2023, 16(2), 827.

[40] Wu, C., Li, J., Liu, W., He, Y. and Nourmohammadi, S. Short-term electricity demand forecasting using a hybrid ANFIS–ELM network optimised by an improved parasitism–predation algorithm. Applied Energy, 2023, 345, 121316.